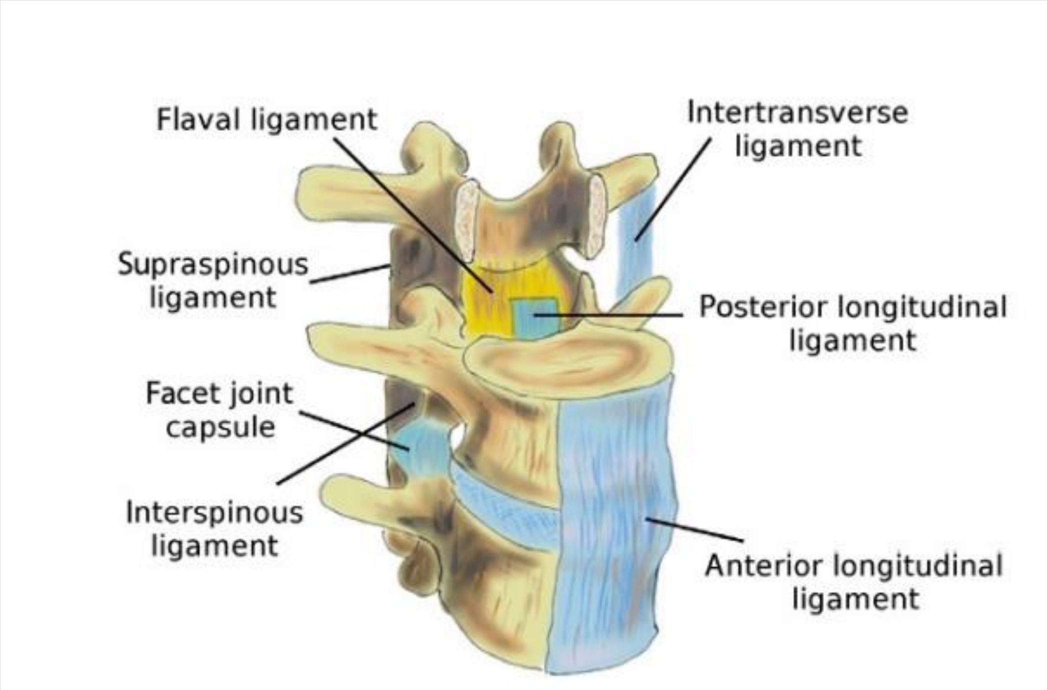
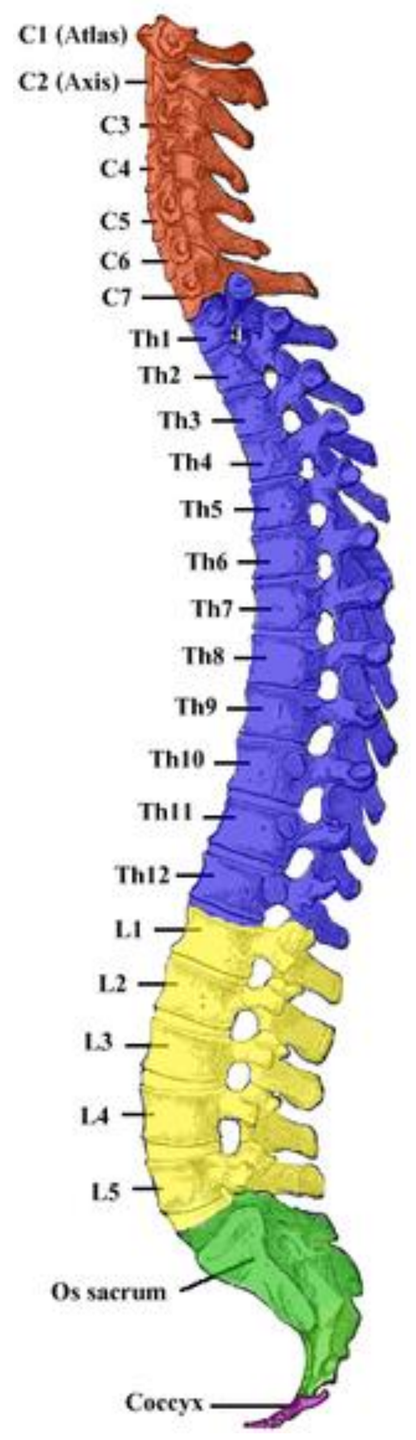
3D Point-of-Interest Prediction on Human Vertebrae Using Spine Segmentations

Daniel-Jordi Regenbrecht

Background and Motivation



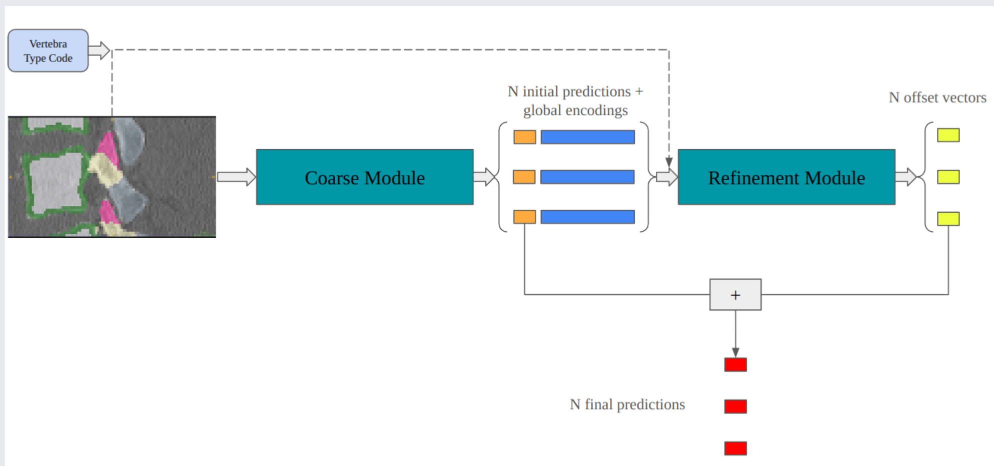
- Lower Back Pain is wide-spread yet poorly understood
- Patient-Specific simulations can help!
- Ligaments transfer the muscular forces to the vertebrae
- Thus, accurate localization of attachment points is vital
- Manual annotation is costly and suffers from intra- and inter-operator-bias
- Previous works have seen great success training on masks or other shape representations



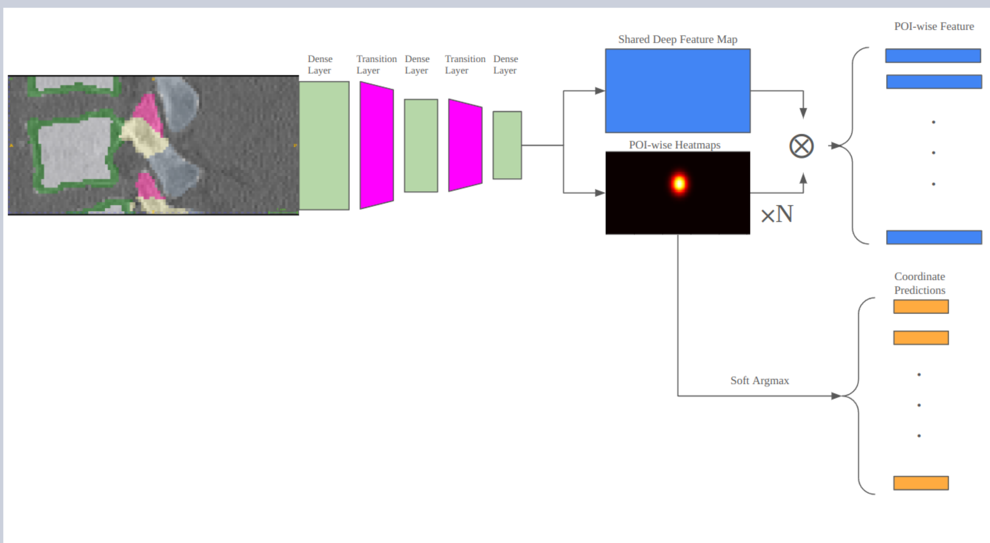
Method

- A baseline based on registration was already available and analyzed
- Additionally, a neural network was designed, implemented and compared to the baseline
- The architecture is two-tiered
  - A "coarse module" implements heatmap regression
  - A "refinement module" employs a transformer to relate initial predictions to each other and refine them

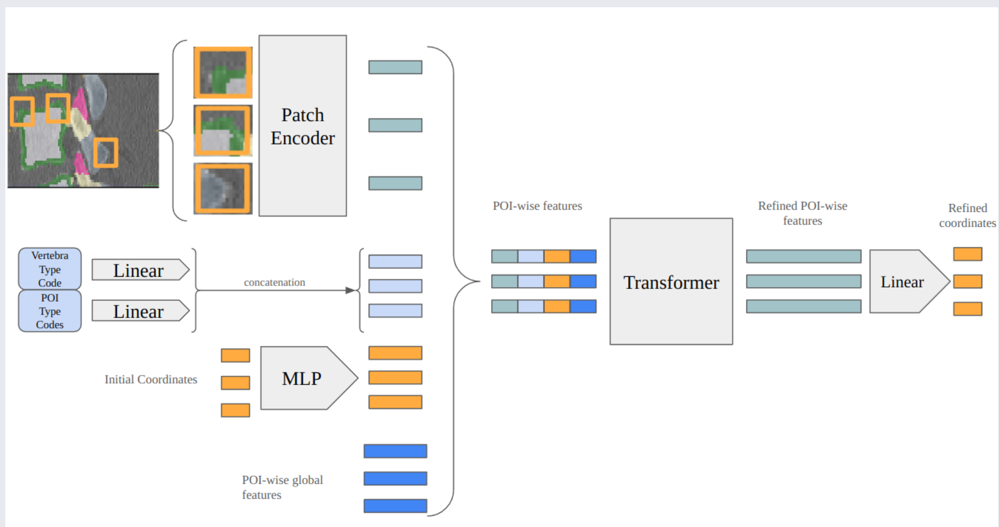
- Overview: Coarse module and refinement module



- Coarse Module: DenseNet-based Heatmap and Feature Map Predictions
- Heatmap provides Coarse Coordinates and is used to extract POI-wise visual features



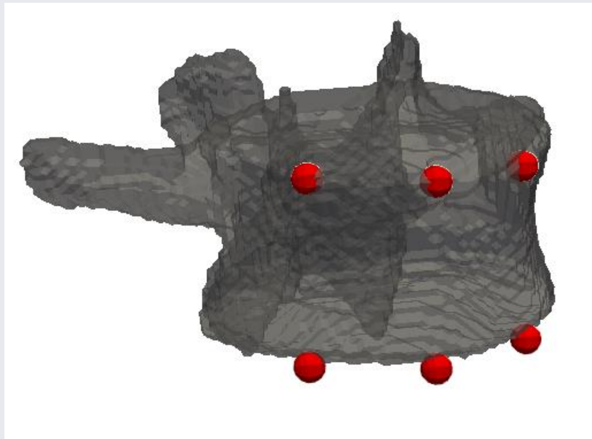
- Refinement Module: Inspired by Visual Transformer
- Relates POI-wise features to each other
- Predicts an Offset to Coarse Predictions



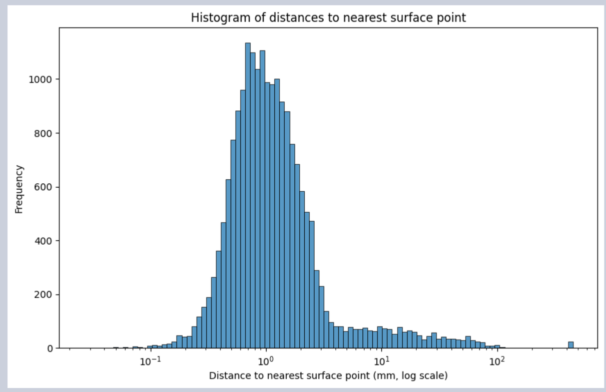
Dataset Composition, Analysis and Cleaning

36 subjects, for each of which there is:
- CT scan (3D Grayscale Values)
- Instance Segmentation Mask for Vertebrae
- Semantic Segmentation Mask for Subregions
- Coordinates of 23 ligament attachment points per vertebra

- Ligaments attachment sites are often better described as areas rather than points
- The dataset thus contains the outer points as well as mid-points for larger attachment sites



Dataset was cleaned based on:
1. Distance to Vertebra Surface
2. Correct spatial arrangement
3. POI lies on correct vertebra subregion



The dataset was split into 23/4/4 subjects for train/val/test. Individual vertebrae were cut out from the semantic segmentation mask and used as model input.

Results

Registration vs Model

Split	Prediction Type	Mean Dist.	Median Dist.	MSE	Accuracy (2mm)
Val	Registration	3.04	2.45	**4.15**	0.38
	Neural Network (NN) Model	**2.17**	**1.73**	7.52	**0.51**
Test	Registration	2.96	2.45	**3.61**	0.36
	NN Model	**2.19**	**2.00**	7.36	**0.50**

Comparison of Coarse and Fine Predictions

Split	Prediction Type	Mean Dist.	Median Dist.	MSE	Accuracy (2mm)
Val	Coarse	2.38	2.24	8.09	0.42
	Fine	**2.17**	**1.73**	**7.52**	**0.51**
Test	Coarse	2.32	2.24	7.75	0.43
	Fine	**2.19**	**2.00**	**7.36**	**0.50**

Qualitative Analysis

- Analyzing the worst cases shows ambiguous or wrong annotations in ground truth (prediction left, ground truth middle, reference image right)

Training on a Subset of POIs

- The POIs in the sagittal plane seem to be much less affected by these deviations. The model was evaluated and re-trained on sagittal POIs only (left table)
- In an additional experiment, POIs with too high disagreement between prediction and ground truth were eliminated during CV (right table)

Split	Trained On	Mean Dist.	Median Dist.	MSE	Accuracy (2mm)
Val	Sagittal	**1.57**	1.41	**3.61**	**0.67**
	Full	1.68	1.41	4.21	0.63
Test	Sagittal	**1.59**	1.41	**3.77**	**0.66**
	Full	1.77	1.41	4.55	0.59

Split	Prediction Type	Mean Dist.	Median Dist.	MSE	Accuracy (2mm)
Val	Coarse	2.08	2.24	5.38	0.44
	Fine	**1.66**	**1.41**	**3.72**	**0.62**
Test	Coarse	2.84	2.45	13.00	0.35
	Fine	**2.63**	**2.24**	**12.02**	**0.44**

Conclusion / Main Findings

- The designed architecture outperforms the baseline and achieves competitive performance compared to similar tasks (cephalometry)
- Eliminating "problematic" POIs shows significant improvements in performance
- This leads to the assumption that data quality is the biggest issues. Future works may
  - Use active learning to improve the ground truth
  - Require a more rigorous definition for the location of some POIs
  - Use synthetic data for more robustness

Klinikum rechts der Isar
Technische Universität München